AUTOMATIC FILTERING OF FAR OUTLIERS IN MULTIBEAM ECHO SOUNDING DATASET USING ROBUST DETECTION ALGORITHMS

Mohd Razali Mahmud[1] and Othman Mohd Yusof[2]

[1]Institute for Geospatial Science and Technology, Faculty of Geoinformation Science and Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia
[2]Department of Surveying Science and Geomatics, Faculty of Architecture, Planning and Surveying Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia
Tel: [1] 607-5530827
Fax: [1] 607-5530827, [2] 607-5566163
E-mail: [1] razali@fksg.utm, [2] omy2006@yahoo.co.uk

Abstract

Bathymetric data collections using multibeam echo sounder (MBES) have led to increasing data rates and densities. While it is really advantage having full coverage of seabed, data management is the utmost aspect to establish. In this data collection method, part of the dataset contains erroneous data, as measurements are always associated with uncertainties. The critical task for hydrographic surveyor is to make decision on which data can be accepted as good data and the remaining data will be considered as outliers. As there is no ground truth available for the MBES data to compare with, the best solution to address this problem is by using statistical outliers elimination. In order to obtain meaningful results when statistical tools are in used, the dataset should be in a Gaussian distribution. To ensure that the dataset in a bell-shaped curve characteristic, the far outliers must be eliminated prior to any processing. This certainly needs further considerations on characteristics of the erroneous data. Thus, a post-processing program was developed to detect and discard the MBES far outliers based on behaviours of propagated beam in the multibeam sonar system. The entire data have to go through a series of far outliers screening. A remarkable result can be achieved by filtering these far outliers using automatic detection mode. This paper elaborates the techniques used for the detection and elimination of the far outliers in the MBES dataset, known as robust detection algorithms. It also explains on the filtering sequences used and results produced by the developed program.

Keywords: *MBES, far outliers, robust detection*

1.0 MBES DATASET PROCESSING

Rapid seafloor mapping is feasible with the existence of multibeam sonar systems. The multibeam echo sounder (MBES) has improved survey technique to determine bathymetry of seabed with the ability to provide a high resolution data collection and full bottom coverage. Data being collected during survey operations can easily be more that 10^6 depth measurements per hour leading to over 10^7 per survey (Capena *et al*., 1999). This case is even worse in shallow water area, where high density of data is produced from a multibeam survey in a more spatially dense and more uniformly distributed over the survey area.

These raw echo records from multibeam sonar usually contain of some erroneous measurements due to reflections upon fish, air bubbles and suspended debris, thus will affect the quality of contours presentation. It is therefore necessary to check for erroneous data called outlier and remove them accordingly. The integrity of acquired data during the sounding operation could be validated if a ground truth validation could be performed. However this is not the case since a systematic checking on the grounds of multibeam data is impossible (Mori, 2003).

In a more traditional manual outliers detection technique using a computer graphic editing tools, a great deal of time is needed and certainly would not a cost effective. In MBES data cleaning process, eliminating for outliers and at the same time not to remove good data is quite a complex assignment. This procedure will rely on subjective decisions make by an experienced operator to decide whether the investigated points are true image of the seabed or merely outliers. The results turn to be inconsistent and not repeatable with different or even the same operators.

Due to high data density, time taken for MBES data processing would be longer especially by the computer graphic line-by-line inspections (Cronin *et al*., 2003). Generally, the manual visual editing process has become very time consuming especially in shallow water surveyed areas. At this junction, it is strongly recommended that to overcome these shortcomings of the manual human interventions to data cleaning such as repeatability, consistency, time constraint, and operator error, automatic outlier detection and elimination needs to be developed.

For the above purpose, a thorough study has been carried out in order to identify the outlier detection criteria based on the MBES beam characteristics during it being propagated through the seawater. Results from this study can be concluded that the outliers can be categorized into two, namely far outlier and near outlier.

This paper will elaborate on the established criteria used to detect and eliminate the far outliers in MBES dataset. It also explains on why the far outliers need to be discarded prior to the process of identifying near outliers and what are the damages can the far outliers contribute to the data cleaning process. Various algorithms and techniques are adapted and transformed into the automatic detection and elimination of far outliers via programming language Visual Basic version 6.0. Part of the results yielded through the developed programs will be discussed towards the end of this paper.

2.0 CONCEPT OF FAR OUTLIERS ELIMINATION

As no measurement is perfect and always related to uncertainties, the whole idea of the programs is data assessment and uncertainty management to the MBES data. Hydrographic uncertainty can be calculated, represented and modeled. Uncertainty management involves the design of program algorithms, programming and the evaluation of results. Three different types of uncertainties in MBES dataset can be categorized as accidental, systematic or random.

As each type must be dealt with differently, this research is mainly focused for “data cleaning”, the term used to describe methods specifically deal with “accidental” and random uncertainties. Systematic uncertainties in MBES data set are the uncertainties that are usually discriminated and omitted during the data collection such caused by heave, roll, yaw, squat of the survey vessel, inconsistency of the speed of sound in water column, ray bending, latency, tidal reduction, offset between the MBES transducer and GPS antenna, etc. This is possible by setting relevant parameters in the navigation software prior to data collection.

The other two uncertainty categories have to go through series of screening. Due to no ground truth of the surveyed seabed area available to compare with MBES data, the most appropriate standard method can be applied is statistical elimination tools. These tools can be used to estimate the seabed surface. However there is a limitation when applying these statistical tools; a statistical phenomenon called *masking effect*. This phenomenon could cause the standard deviation and mean values of the sample to be over estimated. Hence the outliers detection process failed. The simplest case to explain this phenomenon is when a several outliers in the same sample are so far from the sample’s mean value. In this situation, the sample’s standard deviation will greatly inflate thus caused the statistical elimination test insensitive. As a result, all outliers in that sample are undetected and they are considered part of the true depth.

To overcome this phenomenon and in order to obtain meaningful results if the statistical data cleaning technique is to be used, the MBES data set needs to be in a Gaussian distribution pattern.

That would means that, if all accidental and systematic uncertainties have been removed, the remaining datasets contained of purely random uncertainties (Hughes Clarke,1999). The random uncertainties tend to have a special statistical character, known as a Gaussian distribution; an approximate model of reality. Based on the above statement, in this research the outliers are categorised into two groups, called far outliers and near outliers. The far outliers (jointly influential outliers) magnitudes are between 6σ and 9σ whereby the near outliers (random outliers) magnitudes are within 3σ and 6σ (Hekimoglu, 1999).

The far outliers will distort results and defeat the purpose of any statistical analysis and mathematical models to a great extent. The far outliers must be eliminated in the first stage in order to ensure that the remaining datasets are in the bell-shaped curve character. Within the remaining datasets are the near outliers and can be assumed that the near outliers are random uncertainties thus would be detected by the statistical elimination tools. For that reason, the robust filtering is introduced prior to any depth estimation using mathematical models and statistical analysis.

3.0 FAR OUTLIERS IN MBES DATASET

As the MBES generates and continually transmits numerous sonar beams in a swathe or fan-shaped signal pattern, chances for outliers occurrence are unavoidable. The main concern is how to identify and discard the far outliers. A single displacement in a group of observations is the example of the far outlier. It can be also referred to point that deviate so obvious from its neigbourhood point trends or normal anomalies. It comes from different population from the sample being studied. In this MBES sounding data, these far outliers do not represent the overall trend of seabed surface. If these far outliers are unnoticeable and not being eliminated from the dataset, the depth estimation using statistical models will not yield accurate results.

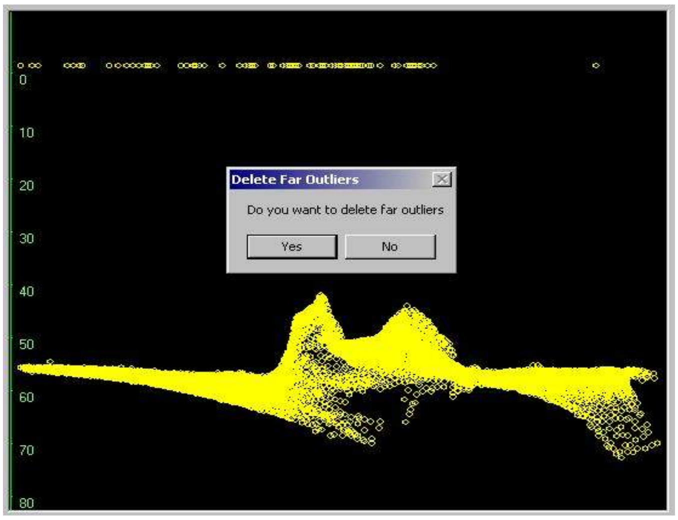

Figure 3.1 Example of far outliers located at less than zero value depth

Figure 3.1 above clearly shows how the far outliers behave in the MBES datasets. Some of them are not clearly visible in the computer graphic view and needs further investigation techniques to be identified. The techniques used robust algorithms that were fully utilized in the post-processed programs.

The concept of automatic detection and elimination the MBES far outliers were based on MBES outliers characteristics. The designed programs mainly focus on MBES post-processing technique. Various algorithms published in journals, websites, articles and manuals were thoroughly studied before the final algorithms can be concluded. Some useful parts of these published algorithms were applied in a customized manner and various approaches were combined in the final stage of algorithms.

4.0 MBES FAR OUTLIERS DETECTION AND ELIMINATION USING ROBUST ALGORITHMS

A series of filtering under robust algorithms was designed and utilised in the programs with intention to detect and eliminate the far outliers. Robust algorithms mean the algorithms are able to find an approximation of the real value even if a high number of outliers in the same sample are placed far from the real value (Capena *et. al*, 1999).

Under the MBES robust filtering algorithms, the following criteria are used to facilitate the detection and elimination of MBES far outliers:

a. Quick View
b. Depth gating
c. Outer beam limit
d. Across-track line anomalies
e. Along-track line anomalies

These various filtering algorithms are selectable by the user during processing. One can select all filtering or skip certain filtering based on the user requirements. The proposed sequences for the detection and removing outliers are explained accordingly in the following discussions.

4.1 Quick View

The easiest way to visualize the far outliers is by running the menu called Quick View. The Quick View menu is designed to indicate the far outliers by displaying vertical view of the swathes. The outliers are significantly recognised as they are located at depths that are far from the rest of the measured depths. The outlier points are normally scattered at certain level of depth with no continuity in seabed anomalies with their neighbourhood. With this trend, user can easily identify the far outliers, thus eliminate these points.

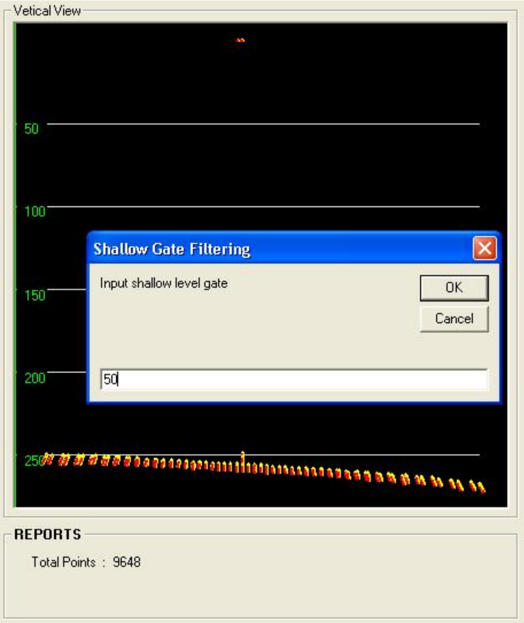

Figure 4.1: Far outliers detection by quick view algorithm

4.2 Depth Gating

One of the ways to eliminate far outliers is by implementing depth gating technique. In this technique, knowledge of overall depth of surveyed area is essential. In other words, the minimum and maximum depth values of the surveyed area must be pre determined before carrying out the processing. With these gating values, any point exceeding the maximum depth limit or less than minimum depth limit is trapped and considered as an outlier, thus will be deleted from the data set. Every single point is tested against these limit values specified by users. Most of the far outliers are detected by this technique.

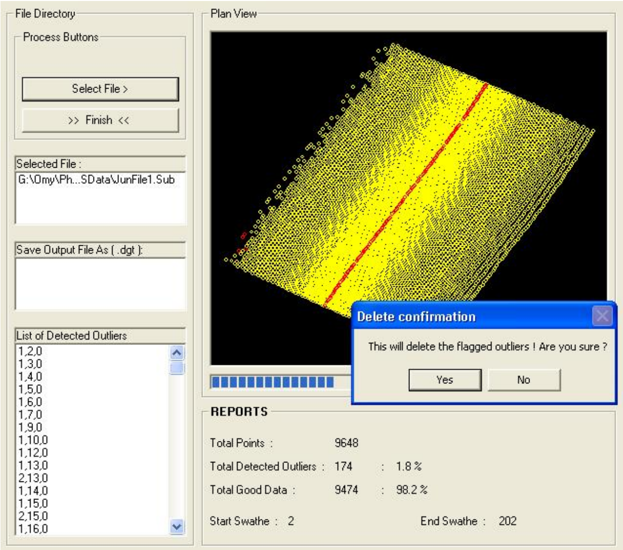

Figure 4.2: Far outliers detection by depth gate algorithm

4.3 Outer Beam Limit

Another direct way to detect far outliers is by limiting the outer beam value on both sides of the swathe direction. Most of the outer beam sectors experienced lower signal-to-noise ratio, created noisier bottom detection thus more occurrence of far outliers are detected in these outer sectors of the swathe. This phenomenon had been proved by de Moustier (2003). In his studies proved that outlier occurrences were more rapid at the most outer beams, especially for the beams generated at more than 65 degrees, measured from the nadir beam. The default beam limitation in this program is 60 degrees but alternatively, user can select other figures between 0-90 degrees as appropriate. Points that exceed the user's limits are flagged as outliers and will be deleted.

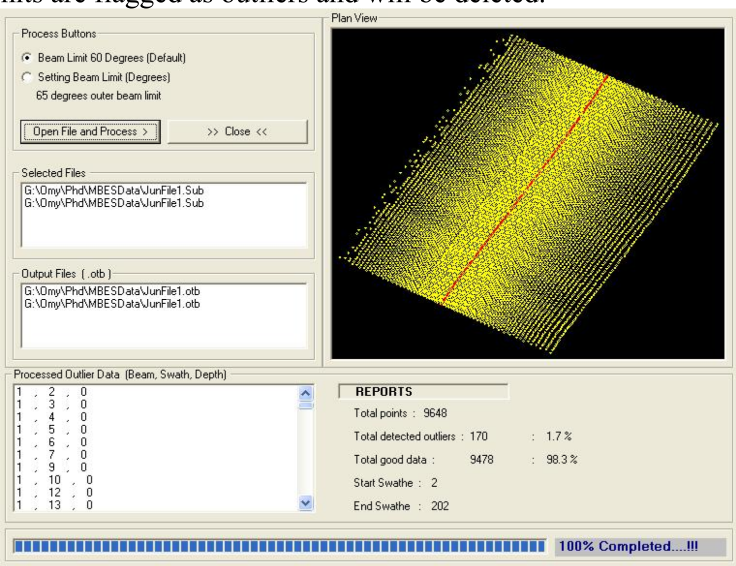

Figure 4.3: Far outliers detection by outer beam limit algorithm

4.4 Across-Track Line Anomalies

General trend of the seabed is said to be continuously smooth and very rare it drastically turns to be very sharp edge or spike, which represented a true seabed feature. Taking advantages of this seabed surface anomaly, a 25 degrees slope limit is specified as a test condition. This filtering tests on every consecutive point in across-track direction. Gardner *et al.*, 1998 in his study concluded that the across-track slope anomalies if exceed 25 degrees will be considered as an outlier. In this algorithm, every point in the same swathe is investigated for its slope with respect to adjacent points. Lines produced between the investigated point and its two adjacent points within the same swathe if more than 25 degrees slope will be considered as an outlier.

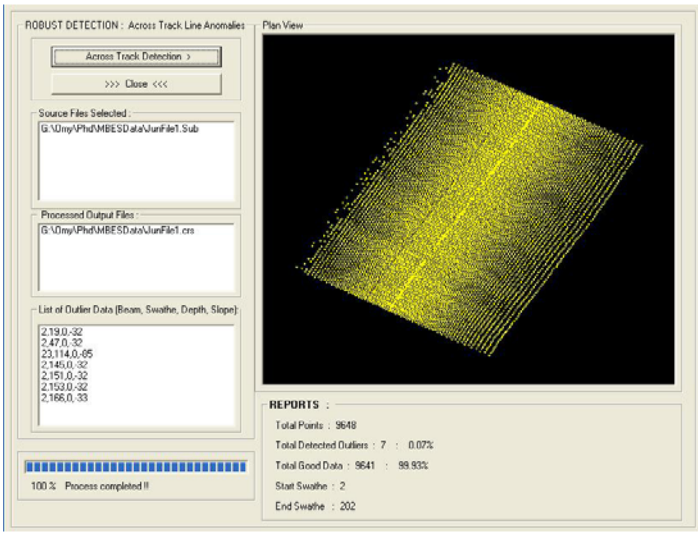

Figure 4.4: Far outliers detection by across-track line anomalies

4.5 Along-Track Line Anomalies

The along-track line anomalies investigate the slope angle anomalies produced between neighbourhood points along the survey vessel track direction as oppose to the above across-track line anomalies. The investigated point will be considered as far outlier if the slope angle generated between the point and its adjacent points more than 55 degrees and the point will be deleted. Both slope angles are in opposite signs before the intermediate point can be said as outlier. The 55 degrees value is accepted as a limit based on study carried out by Gardner *et al.*, 1998.

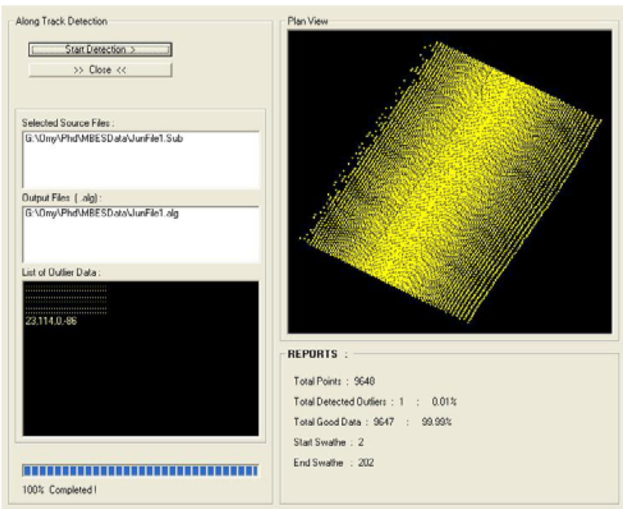

Figure 4.5: Far outliers detected by along-track line anomalies

5.0 RESULTS AND CONCLUSION

Robust algorithms are straightforward applications and quite simple algorithms but yet produce very significant results in tracing and discarding the far outliers in the MBES dataset. A lot of iterative procedures involved in the developed programs. Although they look very simple process but when dealing with high density of MBES data, human intervention and visual investigation technique using a computer graphic editing tools is a tremendous task. A slow line-by-line manual inspection technique is out of date and later on would create up a huge back log of unprocessed data. The human visual analysis has a limited ability and failed to process data quickly and efficiently.

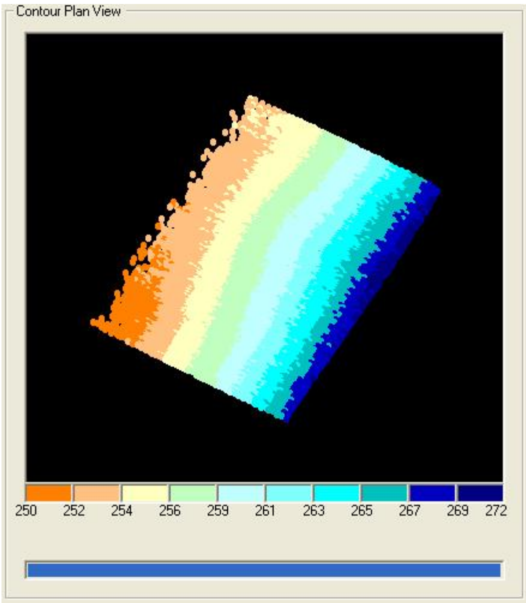

Figure 5.1: Result represents coloured contour plan after MBES far outliers data cleaning

With the automatic outliers detection and elimination, time taken for MBES data cleaning are shorter compared to a long established manual inspection thus give a great impact on the processing techniques.

User can use any sequence order of the five algorithms designed in the programs as explained in the Section 4. Although many other far outlier detection algorithms can be used to detect and discard the far outliers, these five criteria have served perfectly to the expectation standard and fulfill the project objectives. Each algorithm has its own flow and style in detecting the far outliers, therefore the total number of detected outliers reported even though using the same MBES data set would not be the same.

With this automatic data cleaning platform, it lays the remaining dataset with the Gaussian character and should be treated as random uncertainty. This condition permits the use of statistical elimination tools called cross-validation technique in the next detection procedure for the near outliers.

REFERENCES

1. Capena, G., Bergem, O. and Pace, N.G. (1999). Trismap: A Fast way to Deal With New Multibeam Sonar Data. Sea Technology, Vol. 40, No. 10, October 1999, pp. 49-52.

2. Cronin, D., Broadus, M., Byrne, S., Simmons, W. and Gee, L. (2003). Hydrographic Work Flow – From Planning to Products. The Hydrographic Society of America, U.S. Hydro 2003 Conference, Mississippi: March 24-27, 2003.

3. de Moustier, C. (2003). Field Evaluation of Sounding Accuracy in Deep Water Multibeam Swathe Bathymetry. IEEE Webmaster.
a. http://www.ieee.org/organizations/society/oes/html/fall02/Field_eval.html;
b. site visited: September, 2003.

4. Gardner, J. V., Mayer, L. A. and Hughes Clarke, J. E. (1998). Cruise report, RV inland surveyor Cruise is-98; The Bathymetry of Lake Tahoe, California-Nevada. Open-File Report 98-509, USGS.

5. Hekimoglu, S (1999). Robustifying Conventional Outlier Detection Procedures. Journal Of Surveying Engineering, May 1999, pp. 69-86

6. Hughes Clarke, J. E. (1999). Provisional Swathe Sonar Survey Specifications. National Topographic and Hydrographic Authority, Land Information New Zealand, TH Technical Report #2, August 1999.

7. Mori, A. (2003). Seamless Multibeam Data Analysis and Management. Hydro INTERNATIONAL, Vol. 7, No. 3, April 2003, pp. 43-45.